**News Article Classification (Fake vs Real)**

**Introduction**

In an era where digital content spreads faster than ever, the ability to identify misinformation is critical. This project focuses on building a machine learning model that can classify news articles as either fake or real. By combining basic NLP techniques with a clean deployment interface, it offers a reliable, lightweight solution to support media literacy and online credibility.

**Abstract**

The model was developed using a labelled dataset of news articles, employing natural language processing for data preparation and classical machine learning algorithms for classification. Text data was cleaned, tokenized, and transformed into feature vectors using TF-IDF. A Logistic Regression model was trained to distinguish patterns associated with genuine versus deceptive content. The final tool is deployed through a user-friendly Streamlit interface, allowing users to test any article and receive instant feedback on its credibility.

**Tools & Technologies Used**

* **Language**: Python
* **Libraries**: NLTK, Scikit-learn, Pandas, NumPy
* **Model**: Logistic Regression / Naive Bayes
* **Feature Engineering**: TF-IDF Vectorizer
* **Deployment Platform**: Streamlit

**Development Process**

1. **Dataset Selection**: Used a public Kaggle dataset with articles labelled as fake or real.
2. **Text Preprocessing**:
   * Cleaned raw text by removing punctuation and stop words.
   * Tokenized the text and normalized it for consistent formatting.
3. **Feature Extraction**:
   * Applied TF-IDF to convert words into weighted numerical vectors.
4. **Model Training & Evaluation**:
   * Trained both Logistic Regression and Naive Bayes classifiers.
   * Evaluated performance using accuracy, precision, recall, and F1-score.
5. **User Interface**:
   * Developed a Streamlit-based web interface where users can input custom news articles.
   * Predictions are presented along with basic interpretive feedback.
6. **Deployment**:

Model hosted locally through Streamlit for demonstration purposes.